Pendeteksian Relasi Antar Makna Pada Wordnet Bahasa Indonesia

Muhamad Iffandi Pribadi
Teknik Informatika – Universitas Komputer Indonesia
Jalan Dipatiukur No 114-116 Bandung 40132
Bandung, Indonesia
iffandipribadi@gmail.com

Ken Kinanti Purnamasari
Teknik Informatika – Universitas Komputer Indonesia
Jalan Dipatiukur No 114-116 Bandung 40132
Bandung, Indonesia
ken.kinanti@email.unikom.ac.id

Abstrak - Perkembangan WordNet saat ini sudah dicoba diaplikasikan di berbagai negara seperti bahasa-bahasa Arab, Spanyol, Perancis, Belanda dll. Meskipun Wordnet Pertama diciptakan dengan cara manual, pengembangan selanjutnya dilakukan dengan teknik otomatis dan semi otomatis untuk membuat Multilingual WordNet. Metode tersebut secara umum dibagi dua cara yaitu Merge Approach dan Expand approach. Merge Approach adalah metode yang digunakan Princeston University dalam membuat WordNet, selain menghabiskan banyak waktu, merge approach juga sangat mahal untuk dibangun, karena harus melibatkan lexicographer untuk membuat synset. Lalu pendekatan selanjutnya adalah Expand Approach, berbeda dengan merge approach, expand approach mentranslasi synset yang ada di Princeston WordNet(PWN) ke target Bahasa dan mengambil semua relasi yang ada di PWN. Expand approach membutuhkan validasi manual agar informasi yang dihasilkan tidak ambigu.

Pada penelitian sebelumnya mengenai pengembangan wordnet bahasa indonesia data yang dihasilkan dari kategori noun, verb, adj, dan adverb hasil yang didapat untuk Synset adalah sebesar 40774 dari total 117791 Synset dan Unique Strings dihasilkan 23964 dari 155467. Pada proses translasi kategori noun yang berdampak besar untuk presentase hasil uji hanya didapat 15,4 % untuk Unique Strings. Masalah dalam penelitian sebelumnya yaitu tidak lengkapnya kata terjemahan dari hasil translasi menggunakan MRD Cambridge Dictionary dikarenakan bentuk kata dari noun kebanyakan adalah istilah kata benda yang bersifat regional dalam Bahasa Inggris. Banyak istilah-istilah Medis, Kimia, dan Istilah untuk kamus khusus lainya yang tidak dapat diterjemahkan dan kata majemuk dalam Bahasa inggris sangat sedikit terjemahan Bahasa Indonesia nya.

Pada tahap pertama, hasil ekstraksi Unique Strings sebanyak 100% dari total data awal 155467 dan Synset sebanyak 100% dari total data awal sebanyak 117791. Pada tahap kedua pada proses translasi didapatkan Unique Strings sebanyak 15.4% dan Synset sebanyak 34.6%. Hal tersebut dikarenakan kualitas dari MRD Sendiri. Banyak lema atau istilah lokal dalam bahasa Inggris yang tidak ada dalam bahasa Indonesia.

Kata Kunci : wordnet, wordnet Bahasa Indonesia, Expand approach, Automatic Translation.

I. PENDAHULUAN

Perkembangan WordNet saat ini sudah dicoba diaplikasikan di berbagai negara seperti bahasa-bahasa Arab, Spanyol, Perancis, Belanda dll. Meskipun Wordnet Pertama diciptakan dengan cara manual, pengembangan selanjutnya dilakukan dengan teknik otomatis dan semi otomatis untuk membuat Multilingual WordNet. Metode tersebut secara umum dibagi dua cara yaitu *Merge Approach* dan *Expand approach*. *Merge Approach* adalah metode yang digunakan Princeston University dalam membuat WordNet, selain menghabiskan banyak waktu, *merge approach* juga sangat mahal untuk dibangun, karena harus melibatkan *lexicographer* untuk membuat *synset*. Lalu pendekatan selanjutnya adalah *Expand Approach*, berbeda dengan *merge approach, expand approach* mentranslasi synset yang ada di Princeston WordNet(PWN) ke target Bahasa dan mengambil semua relasi yang ada di PWN. *Expand approach* membutuhkan validasi manual agar informasi yang dihasilkan tidak ambigu.

Pada penelitian sebelumnya mengenai pengembangan wordnet bahasa indonesia data yang dihasilkan dari kategori *noun*, *verb*, *adj*, dan *adverb* hasil yang didapat untuk *Synset* adalah sebesar 40774 dari total 117791 *Synset* dan *Unique Strings* dihasilkan 23964 dari 155467. Pada proses translasi kategori noun yang berdampak besar untuk presentase hasil uji hanya didapat 15,4 % untuk *Unique Strings*. Masalah dalam penelitian sebelumnya yaitu tidak lengkapnya kata terjemahan dari hasil translasi menggunakan *MRD Cambridge Dictionary* dikarenakan bentuk kata dari noun kebanyakan adalah istilah kata benda yang bersifat regional dalam Bahasa Inggris. Banyak istilah-istilah Medis, Kimia, dan Istilah untuk kamus khusus lainya yang tidak dapat diterjemahkan dan kata majemuk dalam Bahasa inggris sangat sedikit terjemahan Bahasa Indonesia nya [1]

Berdasarkan masalah di atas maka penelitian ini akan mengembangkan penelitian sebelumnya dengan membangun aplikasi wordnet Bahasa Indonesia dengan melengkapi elemen relasi antar makna seperti antonim, hipernim-hiponim, holonim-meronim dan meningkatkan hasil *Unique Strings* dan *Synset* dengan menggunakan *expand approach.* dengan menggunakan *Oxford Dictionary*, KBBI dan Thesaurus**,** pada penelitian ini memakai *resource Oxford Dictionary* karena *Oxford Dictionary* salah satu kamus terlengkap dan data translasi bisa untuk di *Crawling* dari situs *Oxford English Dictionary*. Proses translasi dengan cara mengambil kata baik dari *index.pos* dan *data.pos* untuk kemudian diubah ke dalam target bahasa yaitu bahasa Indonesia sebagai resource target Bahasa, saat ini *Oxford Dictionary* sudah merilis edisi ke-delapan dengan saat ini berjumlah 600.500 lema yang menjadi acuan untuk pembuatan wordnet Bahasa Indonesia. Dengan bantuan MRD Dualingual Inggris-Indonesia dan Indonesia

Inggris atau *machine translation* untuk translasi dari PWN ke Bahasa Indonesia dan dilakukan validasi manual untuk menghindari ambigu dari MRD dan *machine translation* sehingga data yang dihasilkan bisa berkualitas.

II. DATA MASUKAN

Data yang digunakan untuk pengembangan Wordnet Bahasa Indonesia menggunakan Wordnet dari Princeton dan Cambridge Dictionary.

A. *Princeton Wordnet (PWN)*

Database dari PWN mempunyai format sendiri dengan mengelompokan data berdasarkan kelas kata dimana kelas kata ini dikategorikan noun, verb, adjective dan adverb. Format *database* PWN menggunakan encoding ASCII sehingga mudah untuk dibaca oleh manusia dan mesin. *Database* PWN dibagi menjadi 2 tipe yaitu *index.pos* dan data.pos

- File *Index.pos* terdiri dari list kata secara alfabet dengan kelas kata masing-masing *database*. Pada setiap baris list kata tersebut mempunyai elemen yang bernama Synset_offset yang terhubung dengan data.pos. setiap Synset mengandung kata. Kata/lema pada *index.pos* mempunyai format lower case. Pada *index.pos* setiap *database*nya terdiri copyright, version number, license agreement pada awal baris.
- File *data.pos* mempunyai informasi yang mengandung anggota Synset dari *index.pos*. setiap file data.pos diawali dengan copyright notice, version number dan license agreement. Setiap baris data mempunyai informasi yang sudah disusun oleh Leksikografer untuk Synset. Setiap informasi data diawali oleh 8 byte offset atau address suatu Synset[2].

B. *MRD Cambridge Dictionary*

Data masukan berupa kamus dwibahasa Cambridge Dictionary diambil dari website resmi Cambridge Dictionary. Data yang diambil dari hasil crawling adalah kata, subkata dan kelas kata.

Data awal untuk translasi digunakan *Unique Strings* pada database Princeton Wordnet hal itu dikarenakan target kata yang akan diterjemahkan datanya berasal dari Princeton Wordnet. *Unique Strings* kemudian diurutkan berdasarkan abjad dan dikelompokan berdasarkan kelas kata. Setiap kata dijadikan parameter untuk dilakukan pengambilan hasil translasi. Terdapat kesulitan ketika mengambil terjemahan yang berupa kata majemuk atau lebih dari dua kata dikarenakan format link yang ada pada website Cambridge dictionary mempunyai pencarian khusus. Pencarian kata majemuk atau yang lebih dari dua kata diambil kata dasar untuk dilakukan pencarian yang menjadi sulit untuk mendeteksi kata dasar setiap masukan data dari Princeton Wordnet.[4].

III. ISI PENELITIAN

Rumusan masalah menurut latar belakang di atas adalah belum lengkapnya elemen relasi antar makna seperti antonim, hipernim-hiponim, holonim-meronim dan hasil *Unique Strings* dan *Synset* yang belum optimal pada Wordnet Bahasa Indonesia.

A. Wordnet

Wordnet adalah suatu Database leksikal yang terinspirasi teori psikolinguistik tentang leksikal memori pada manusia. Wordnet pertama kali ditemukan dan dikembangkan di Cognitive Science Laboratory Princeton University di bawah arahan Profesor Psikologi George A. Miller pada Tahun 1985 [3]. Perbedaan antara Wordnet dengan kamus bahasa pada umunya adalah Kamus memfokuskan pada kata sedangkan Wordnet memfokuskan kepada makna. Suatu makna pada Wordnet dikelompokan ke dalam set sinonim yang disebut Synset. Wordnet menyediakan singkatan, definisi umum dan mencatat hubungan semantik antara set berbagai sinonim. Tujuanya untuk menghasilkan kombinasi antara kamus dan thesaurus. Wordnet dipakai dalam berbagai aplikasi yang membantu manusia mengolah dan mencerna informasi misalnya di bidang Information Retrieval, Machine Translation, dan Natural Language Processing.

Wordnet mempunyai konsep kamus dengan menggunakan metode *searching* dibanding kamus pada umumnya yang menggunakan susunan alfabet. mencari kata atau lemma dengan susunan alfabet sangat menyita waktu untuk itu Wordnet memudahkan para Leksikografer untuk mengorganisir informasi *Database* leksikal yang dapat dibaca oleh komputer yang tentunya lebih cepat dari manusia. Berbeda dengan kamus, Wordnet fokus pada makna kata jadi, Wordnet dapat disebut juga *online thesaurus*. Saat ini Wordnet yang sudah stabil dan mempunyai *Synset* paling banyak diantara Wordnet lain adalah *Princeton* Wordnet (PWN). PWN juga biasa dijadikan acuan untuk membuat Wordnet di bahasa lain[2].

B. Analisis Masalah

Masalah yang diambil berdasarkan latar belakang yang dibahas sebelumnya adalah dengan mengembangkan dari penelitian sebelumnya. Dikarenakan belum adanya struktur mengenai *Database* Wordnet, relasi antar makna yang bersifat *open source*. dan belum adanya jaringan semantik hasil dari Leksikografer, untuk itu relasi antar makna bahasa Indonesia belum terbentuk dan tersedia. Untuk saat ini pengembangan Wordnet bahasa Indonesia belum dilakukan kembali dan belum adanya sumber daya yang sudah bisa digunakan dan bisa dikembangkan dengan sifat *open source*. Pada penelitian sebelumnya mengenai pengembangan wordnet bahasa indonesia data yang dihasilkan adalah dari kategori *noun*, *verb*, *adj*, dan *adverb* hasil pengujian yang didapat dari pengembangan sebelumnya bisa dilihat pada tabel 1.

Tabel 1. Hasil Pengujian

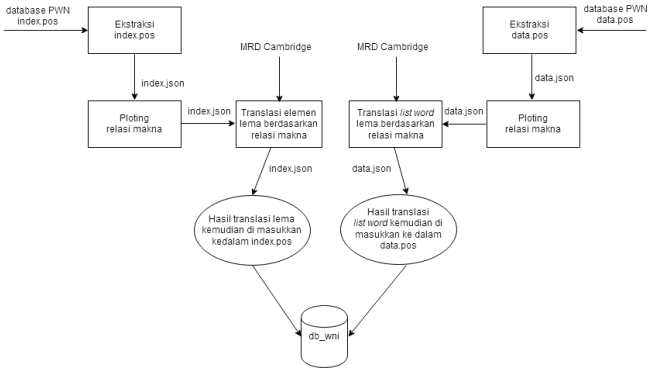
Kategori	Data Awal		Hasil Ekstraksi		Hasil Translasi	
	Unique Strings	***Synset***	***Unique Strings***	***Synset***	***Unique Strings***	***Synset***
Noun	117953	82192	117953	82192	11944	20992
Verb	11540	13789	11540	13789	5283	10290
Adjective	21499	18185	21499	18185	4755	7366
Adverb	4475	3625	4475	3625	1982	2126
TOTAL	**155467**	**117791**	**155467**	**117791**	**23964**	**40774**

Synset di dapatkan hasil sebesar 40774 dari total 117791 *Synset* dan *Unique Strings* dihasilkan 23964 dari 155467 *Unique Strings*. Pada proses translasi kategori noun yang berdampak besar untuk presentase hasil uji hanya didapat sekitar 15,4 % untuk *Unique Strings*. Masalah dalam penelitian sebelumnya yaitu tidak ada nya kata terjemahan dari hasil translasi menggunakan *MRD Cambridge Dictionary* dikarenakan bentuk kata dari noun kebanyakan adalah istilah kata benda yang bersifat regional dalam Bahasa Inggris. Banyak istilah-istilah Medis, Kimia, dan Istilah untuk kamus khusus lainya yang tidak dapat diterjemahkan dan kata majemuk dalam Bahasa inggris sangat sedikit terjemahan Bahasa Indonesia nya. Kebutuhan Wordnet untuk informasi masih belum terstruktur dikarenakan pengembangan dari kamus baru dilakukan sehingga dibutuhkan suatu *resource* untuk mengolah bahasa Indonesia.

C. Analisis Solusi

Analisis solusi merupakan suatu tahapan dimana prosesnya untuk mengidentifikasi setiap masalah dan kebutuhan yang diperlukan untuk membuat percobaan dengan cara lain. Solusi yang diberikan dalam penelitian ini adalah menggunakan Metode *Expand Approach* dengan *Automatic Translation* dengan menerjemahkan setiap lema atau kata yang ada pada *Database Princeton* Wordnet (PWN) dengan cara ekstrak *Database* dari *Princeton* Wordnet dan mengambil setiap relasi semantik yang ada pada PWN. Tahapan pertama yang dilakukan adalah proses pembersihan karakter yang bukan data, seperti *license agreement, number version* dan *copyright notice* yang ada pada awal setiap *Database*. Karena format *Database* PWN berupa *data* dan *index* dan struktur datanya berbeda untuk itu akan dilakukan dua proses yang berbeda ekstrak *Database* berdasarkan *data* dan *index*. Setelah tahap ekstraksi berhasil dilakukan translasi menggunakan *Machine Readable Dictionary* (*MRD*) dari *Cambridge Dictionary* pada setiap lema yang ada pada *Database* PWN. Yang selanjutnya disimpan ke dalam data yang baru. Untuk lebih jelasnya bisa dilihat pada gambar 1.

Pada Analisis solusi untuk sistem pembangunan Wordnet bahasa Indonesia menggunakan *MRD* memiliki tahapan ekstrak *Database* PWN, translasi lema, dan konversi format.



Gambar 1 Gambaran Arsitektur Sistem

Gambar 1 Gambaran arsitektur sistem di atas dijelaskan sebagai berikut.

1. Pada tahap awal *User* memasukan *Database* PWN *index.pos* yang berformat *ASCII* ke dalam sistem lalu dilakukan ekstraksi struktur datanya.
2. Selanjutnya ploting setiap lema pada *index.pos* berdasarkan relasi makna.
3. Tahap selanjutnya dilakukan proses translasi pada elemen pertama yaitu lema. Translasi menggunakan *MRD* dari *Cambridge Dictionary*
4. Lema yang sudah ditranslasi kemudian diurutkan dan disimpan ke dalam ke *Database*.
5. Pada proses ini data masukan adalah *Database* PWN *data.pos* yang berformat *ASCII* ke dalam sistem lalu dilakukan ekstraksi dan pengambilan struktur data, hasil keluaran berformat *.json*
6. Selanjutnya ploting setiap lema pada *data.pos* berdasarkan relasi makna.
7. Pada tahap ini diambil *Wordslist* pada setiap baris dan dilakukan penerjemahan dari bahasa Inggris ke Indonesia dengan bantuan *MRD*.
8. Setelah penerjemahan berhasil kemudian mengambil struktur data yang ada pada PWN dan memasukan hasil translasi ke dalam struktur data yang diambil dan disimpan ke dalam *Database*.

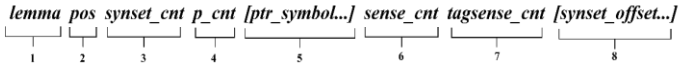
D. Analisis Data Masukan

Analisis data masukan pada sistem yaitu menjelaskan proses data berupa *Database* dari PWN dan data dari *Cambridge Dictionary*. *Database* PWN sendiri mempunyai format khusus dengan mengelompokkan data dengan *data.pos* dan *index.pos* dimana *pos* ini adalah *noun,verb,adjective,* dan *adverb*. Format *Database* PWN menggunakan *encoding ASCII* sehingga mudah untuk dibaca oleh manusia dan mesin.

1. *Index.pos*

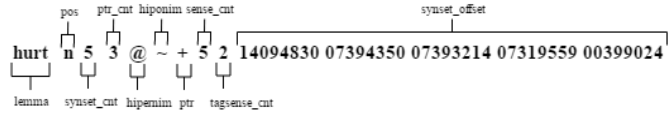
Index file terdiri dari list kata secara alfabet dengan kelas kata masing-masing *Database*. Pada setiap baris list kata tersebut mempunyai elemen yang bernama *Synset_offset* yang terhubung dengan *data.pos*, relasi antar makna di representasian dengan struktur data *[ptr_symbol...]*, setiap *Synset* mengandung kata. Kata/lema pada *index.pos* mempunyai format *lower case*. Pada *index.pos* setiap *Database*nya terdiri *copyright, version number, license*

agreement pada awal baris. Berikut adalah struktur data yang ada pada *index.pos.*



Gambar 2 Data Masukan *index.pos*

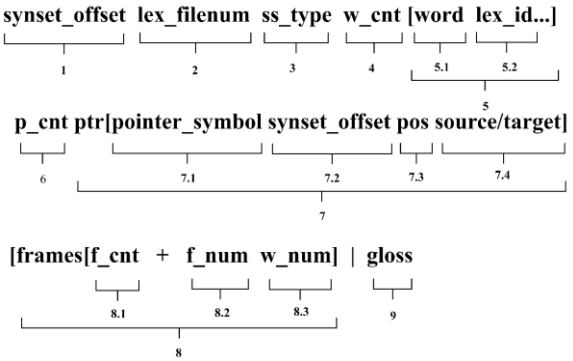
Untuk lebih jelas berikut adalah contoh dari data masukan *index.pos.*



Gambar 3 Contoh Data Masukan *Index.pos*

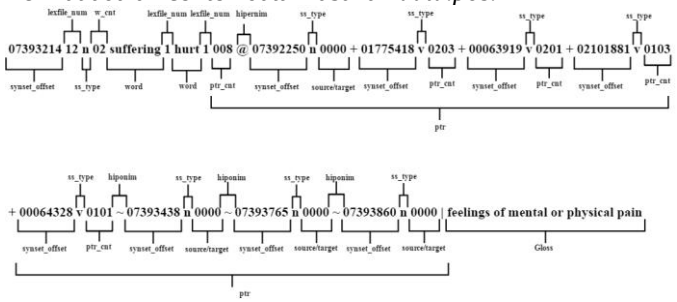
2. *Data.pos*

File *data.pos* mempunyai informasi yang mengandung anggota *Synset* dari *index.pos*. setiap file *data.pos* diawali dengan *copyright notice*, *version number* dan *license agreement*. Setiap baris data mempunyai informasi yang sudah disusun oleh Leksikografer untuk *Synset*. Setiap informasi data diawalai oleh 8 *byte offset* atau *address* suatu *Synset.*



Gambar 4 Data Masukan *Data.pos*

Berikut adalah contoh data masukan *data.pos.*



Gambar 5 Contoh Data Masukan *Data.pos*

E. Analisis Proses

Analisis proses menjelaskan tentang pendekatan sistematis dalam mengidentifikasi setiap permasalahan dan tujuan dalam merancang Wordnet bahasa Indonesia. Analisis proses yang dilakukan adalah ekstraksi dan translasi sebagai proses utama dan konversi data. Karena struktur data pada *data.pos* dan *index.pos* berbeda maka dilakukan proses ekstrak yang berbeda untuk itu dilakukan 2 proses yang berbeda.

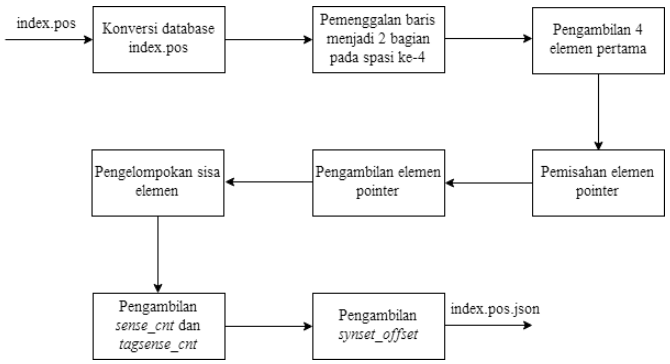
1. Ekstrasi *Index.pos*

Tahap ini berfokus pada pengambilan struktur data pada *index.pos* untuk mendapatkan struktur yang terdiri dari *lemma, pos, Synset_cnt, p_cnt [ptr_simbol...] sense_cnt, tagsense_cnt dan Synset_offset [Synset_offset...]*. data masukan berasal dari *Princeton* Wordnet (PWN) yang berformat *ASCII*. PWN sendiri membagi *Database* nya menjadi empat bagian yang masing-masing dikelompokan berdasarkan kelas kata yaitu *index.noun, index.verb, index.adverb, dan index.adjective* Contoh data masukan yang diambil dari PWN adalah sebagai berikut

Data *Index.pos*

hurt n 5 3 @ ~ + 5 2 14094830 07394350 07393214 07319559 00399024
laugh n 3 4 @ ~ %p + 3 2 07029036 06785996 06687271

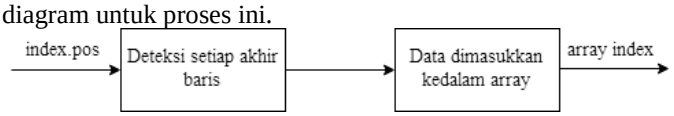
Tahapan tersebut di atas dapat dilihat pada gambar berikut.



Gambar 6 Skema Ekstraksi *Database Index.pos* dari PWN

2. Konversi Database *Index.pos*

Proses memasukan data ke dalam *array* bertujuan untuk memudahkan dalam pemrosesan data. Berikut adalah blok diagram untuk proses ini.



Gambar 7 Blok Diagram Konversi *Database Data.pos*

Pada gambar di atas menunjukkan tahap pertama yaitu memasukan data ke dalam *array* karena setiap data dipisahkan oleh baris baru "/n" maka data akan diproses per-baris dengan memasukan ke dalam *array*. Pada tahap ini *index.pos* harus sudah bersih dari *copyright notice*, *version number* dan *agreement* pada awal kalimat.

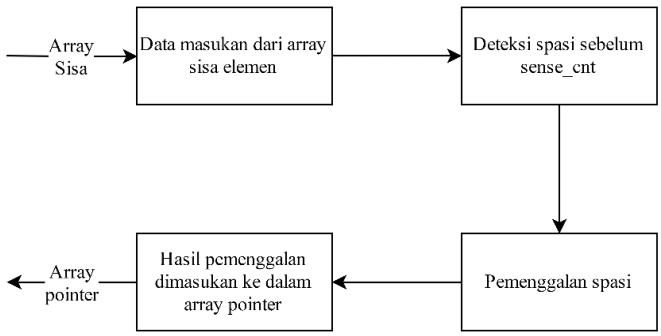
Konversi *Database Index.pos*

Sebelum
hurt n 5 3 @ ~ + 5 2 14094830 07394350 07393214 07319559 00399024 laugh n 3 4 @ ~ %p + 3 2 07029036 06785996 06687271

Sesudah
Array ([0] => hurt n 5 3 @ ~ + 5 2 14094830 07394350 07393214 07319559 00399024 [1] => laugh n 3 4 @ ~ %p + 3 2 07029036 06785996 06687271)

3. Pemisahan Elemen Relasi Makna

Pemisahan elemen Relasi Makna dengan cara Pemenggalan pada bagian kedua menjadi dua bagian,dengan cara membagi dari spasi dekat angka jumlah *Synset*. Untuk lebih jelas nya bisa dilihat pada blok diagram di bawah ini.



Gambar 8 Blok Diagram Pemisahan Elemen Relasi Makna.

Pada contoh Pemisahan elemen Relasi Makna di bawah ini dimaksudkan untuk mengambil elemen *Pointer* (! & \ = + dan lain lain), spasi yang ditandai adalah parameter untuk pemenggalan baris. Hasil pemenggalan masing-masing disimpan di *array* sementara yang nantinya dilakukan proses selanjutnya. Contoh Pemisahan elemen *Pointer* bisa dilihat pada tabel di bawah ini.

Tabel 1 Pemisahan Elemen Relasi Makna

Sebelum
@ ~ %p = + - 3 1 01092370 08146250 07152330 @ ~ ; 2 1 09961754 08421330
Sesudah
@ ~ %p = + - @ ~ ; 3 1 01092370 08146250 07152330 2 1 09961754 08421330

4. Pengambilan Elemen Relasi Makna

Pada proses ini bagian pertama hasil split sebelumnya di penggal berdasarkan spasi dan di masukan ke dalam *array*, *array* tersebut dimasukan kembali ke *array* hasil. Untuk lebih jelasnya bisa dilihat pada blok diagram di bawah ini.

Gambar 9 Pengambilan Elemen *Pointer*

Tahapan yang dilakukan seperti pada Gambar Pengambilan elemen Relasi Makna adalah mengambil *array* pertama dari hasil pemenggalan sebelumnya kemudian di penggal lagi berdasarkan spasi, hasil dari pemenggalan tersebut kemudian di masukan ke dalam *array Pointer* dan selanjutnya dimasukan ke dalam *array* hasil. Untuk lebih jelas tahapanya bisa dilihat pada contoh di bawah ini.

Tabel 2 Pengambilan Elemen *Pointer*

Sebelum
@ ~ + @ ~ %p +
Sesudah
Array ([0] => @ [1] => ~ [2] => +) *Array* ([0] => @ [1] => ~ [2] => %p [3] => +)

IV. HASIL PENGUJIAN

Tahap pengujian sistem bertujuan untuk menemukan kesalahan – kesalahan atau kekurangan – kekurangan pada sistem yang diuji. Pengujian bermaksud untuk mengetahui apakah program kamus kata dan jenis kata yang dibuat sesuai dengan tujuan penelitian.

Tabel 2 Data Awal

Kategori	Data Awal			
	Atonim	Hipernim/Hiponim	Meronim/Holonim	Definisi (Gloss)
Noun	20589	22351	8900	54084
Verb	1786	33640	4320	23433
Adjective	23987	4234	3455	52342
Adverb	2374	3435	1532	52344
Total	**48736**	**63660**	**18207**	**182203**

Pada tahap pertama yaitu ekstraksi *database* dari PWN didapat hasil akurasi 100 %, semua data beserta strukturnya berhasil didapatkan. Pada Tahap kedua dilakukan translasi untuk mendapatkan *Synset* dan *Unique Strings* ke dalam target bahasa, Untuk memperoleh presentase pembangunan sysnset dari hasil translasi oleh *MRD Cambridge Dictionary*. Maka perhitungan yang dilakukan membagi hasil Relasi Makna yang berhasil di translasi dengan jumlah *Synset*s dalam PWN. pada Tahap ke-dua data yang dihasilkan dari kategori *noun*, *verb*, *adj*, dan *adverb* didapat 47996 untuk antonim, 59460 untuk Hipernim/Hiponim, 17357 untuk Meronim/Holonim, dan 138203 untuk Definisi

Tabel 2 Data Hasil

Kategori	Hasil Ekstrasi			
	Antonim	Hipernim/ Hiponim	Meronim/ Holonim	Definisi
Noun	20389	22151	8750	52084
Verb	1746	31640	4120	21433
Adjective	23787	3234	3155	32342
Adverb	2074	2435	1332	32344
TOTAL	**47996**	**59460**	**17357**	**138203**

Pada proses translasi bisa dilihat pada kategori noun yang berdampak besar untuk presentase hasil uji hanya didapat sekitar 10% untuk *Unique Strings* hal tersebut dikarenakan tidak ada nya kata terjemahan dari hasil translasi menggunakan *MRD Cambridge Dictionary* dikarenakan bentuk kata dari noun kebanyakan adalah istilah kata benda yang bersifat regional dalam Bahasa Inggris. Banyak istilah-istilah Medis, Kimia, dan Istilah untuk kamus khusus lainya yang tidak dapat diterjemahkan dan kata majemuk dalam Bahasa inggris sangat sedikit terjemahan Bahasa Indonesia nya.

V. KESIMPULAN DAN SARAN

Sistem yang dibangun dalam hal ini adalah pendeteksian relasi antar makna pada Wordnet bahasa Indonesia telah berhasil dibuat. Dan telah tersedianya suatu database leksikal secara open source. Pada tahap pertama mendapatkan hasil ekstraksi Unique Strings sebanyak 155467 dari total data awal 155467 dan Synset sebanyak 117791 dari total data awal sebanyak 117791. Pada tahap kedua pada proses translasi dihasilkan dari kategori *noun*, *verb*, *adj*, dan *adverb* didapat 47996 untuk antonim, 59460 untuk Hipernim/Hiponim, 17357 untuk Meronim/Holonim, dan 138203 untuk Definisi. Bisa dilihat pada hasil ekstraksi didapatkan hasil ekstraksi sepenuhnya pada tahap pertama, dan pada tahap kedua terjadi penyusutan hasil dikarenakan kualitas dari MRD. Banyak lema seperti kata "batman" yang tidak ada terjemahanya dan banyak istilah-istilah lokal untuk bahasa Inggris yang tidak ada di bahasa Indonesia.

Penelitian selanjutnya adalah menambahkan elemen makna dari KBBI dan penambahan peningkatan *Unique Strings* dan *Synset* dengan metode otomatis atau pembangunan *Synset* secara semi otomatis menggunakan thesaurus dan KBBI. Atau menggunakan *Automatic Translation* beberapa MRD lalu memberikan rangking pada setiap terjemahanya sehingga didapat terjemahan yang paling tepat.

REFERENCES

[1] D. Novriandi, "Development Indonesian Wordnet," 2008.
[2] E. Pianta, L. Bentivogli, and C. Girardi, "MultiWordNet: developing an aligned multilingual database," *Proc. First Int. Conf. Glob. WordNet*, no. 1996, pp. 293–302, 2002.
[3] J. Ilmiah, I. Komputa, E. Volume, and B. Issn, "SEBAGAI SUMBER DAYA NLP BAHASA INDONESIA Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)."
[4] C. Soler, "Extension of the SpanishWordNet," *Gwc 2004 Second Int. Wordnet Conf. Proc.*, pp. 213–219, 2003.